Section 1.4
Defining Data Spread

Curriculum Outcomes	Related Activities	Page in Text
• analyze statistical summaries, draw conclusions, and communicate results about distributions of data	• investigate sets of data with a variety of distributions to develop the concept of dispersion, which includes range and variation	27
• calculate various statistics using appropriate technology, analyze and interpret the displays, and describe the relationships	• investigate sets of data with the same range and mean but different variation, to conclude that variation is independent of range and the location of the data	30
• calculate and apply the mean and standard deviation using technology to determine whether a variation makes a difference	• compute standard deviations of data sets with a variety of variations and relate standard deviation values to the shape of graphs of those data sets	29

Section 1.4 -- Pg.27

Dispersion

A measure of the spread of data including the range of the data and the variation within the set of data.

Note: The "range" gives a good indication of the amount of spread, but is a poor indicator of variation within the data. It also provides no information about the location of the central values.

Standard Deviation

A number that describes the spread within a set of data. It represents the average distance a random piece of data is likely to be located from the mean of the data.

-->In other words:

-If there are lots of values far away from the mean, the standard deviation will be greater than if most of the values are close to the mean.

-Note: The standard deviation does not indicate the location of the data but how the data values within a set relate to each other and to the mean.

Refer to Pg.27-29

Example

The manager of a Christmas-tree farm did a study on the time (in hours) needed to sell scotch pines versus white pines once the trees were cut down. The results were as follows;

	TIMES (hrs)								
Scotch Pines	2.15	7.34	9.14	9.50	10.3	10.5	10.7	12.1	21.7
White Pines	2.28	2.46	3.25	4.25	10.2	13.5	14.4	21.2	21.7

Compare the times for the two types of trees by referring to pages 27 - 29:

1) Both trees have the same *mean* time of 10.4

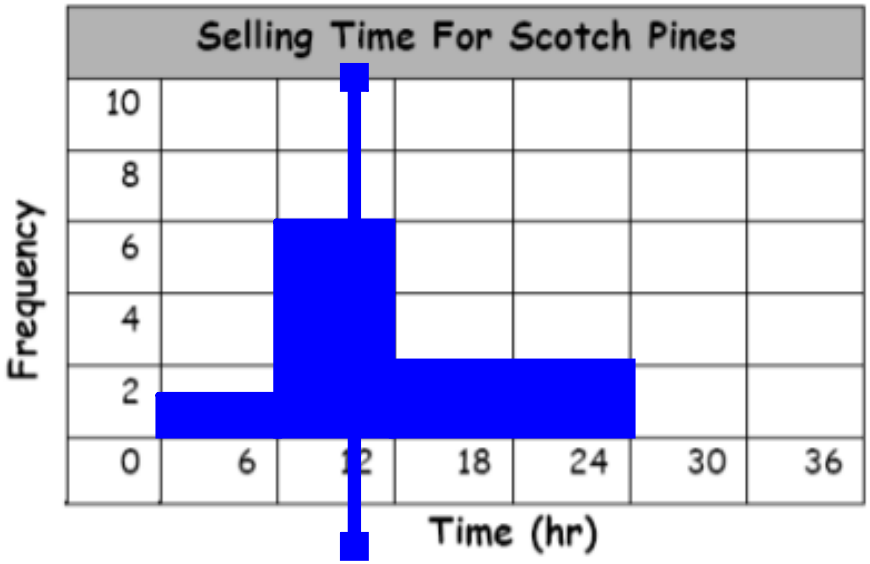
2) Their *ranges* differ only by 0.2hr,

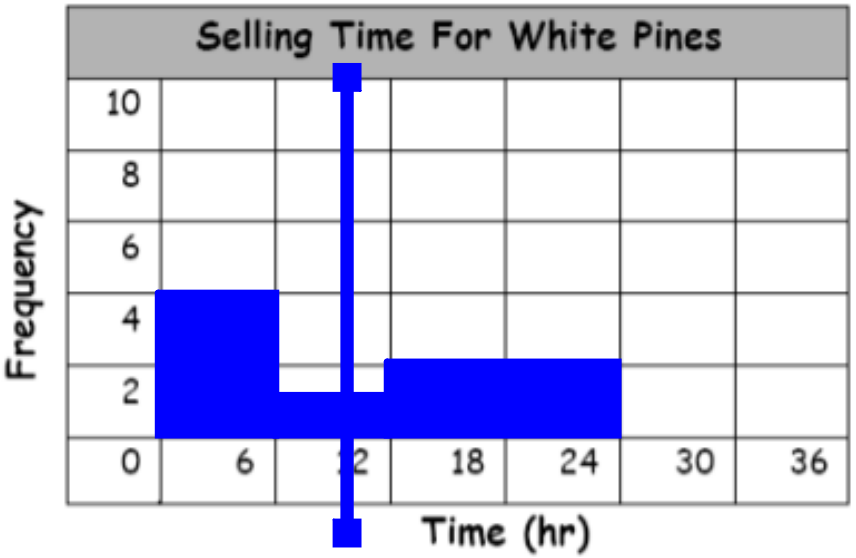
Scotch Pines Range → 19.6 (2.15 to 21.7)

White Pines Range → 19.4 (2.28 to 21.7)

3) Their histograms differ dramatically.

Shade in the appropriate spaces below.





By referring to the *histograms* above, predict which data set would have a **larger** *standard deviation.* →

For a numerical description of the data spread, calculate the Standard Deviation.
First calculate the Standard Deviation for White Pines:

White Pines

Mean $\overline{X}$	Data Value X	Deviations From the Mean $\overline{X} - X$	Squared Deviations
10.4	2.28	10.4 - 2.28 = 8.12	$(8.12)^2 = 65.9$
10.4	2.46	10.4 - 2.46 = 7.94	$(7.94)^2 = 63.0$
10.4	3.25	10.4 - 3.25 = 7.15	$(7.15)^2 = 51.1$
10.4	4.25	10.4 - 4.25 = 6.15	$(6.15)^2 = 37.8$
10.4	10.2	10.4 - 10.2 = 0.2	$(0.2)^2 = 0.04$
10.4	13.5	10.4 - 13.5 = -3.1	$(-3.1)^2 = 9.61$
10.4	14.4	10.4 - 14.4 = -4.0	$(-4.0)^2 = 16.0$
10.4	21.2	10.4 - 21.2 = -10.8	$(-10.8)^2 = 116$
10.4	21.7	10.4 - 21.7 = -11.3	$(-11.3)^2 = 128$

(Sum of Squared Deviations) / (Number of Data Values) = 54.2

[which is the mean of the squared deviations]

The square root of the mean of the squared deviations: = 7.36

Therefore, the **Standard Deviation** (S_x) for White Pines is → **7.36hr**

This means that a random piece of data from the White Pines data set is, on average, 7.36hr away from the mean.

Determine the Standard Deviation for Scotch Pines

Review of Steps for calculating Standard Deviation:

Step #1 = Find the $\overline{X}$ of the data

Step #2 = Calculate the distance each data value is from the mean. Square the difference.

Step #3 = Find the $\overline{X}$ of the sum of the squares.

Step #4 = Find the square root of the mean.

Scotch Pines

Mean $\overline{X}$	Data Value X	Deviations From the Mean $\overline{X} - X$	Squared Deviations $(\overline{X} - X)^2$
10.4		10.4 - _____ =	(_____)^2 =
10.4		10.4 - _____ =	(_____)^2 =
10.4		10.4 - _____ =	(_____)^2 =
10.4		10.4 - _____ =	(_____)^2 =
10.4		10.4 - _____ =	(_____)^2 =
10.4		10.4 - _____ =	(_____)^2 =
10.4		10.4 - _____ =	(_____)^2 =
10.4		10.4 - _____ =	(_____)^2 =
10.4		10.4 - _____ =	(_____)^2 =

(Sum of Squared Deviations) / **(Number of Data Values)** [which is the mean of the squared deviations]	= 23.4
$\sqrt{\frac{(Sum\ of\ Squared\ Deviations)}{(Number\ of\ Data\ Values)}}$	4.84

Therefore, the **Standard Deviation** (S_x) for Scotch Pines is →

This means that a random piece of data taken from the Scotch Pines data set is, on average, _____ hr away from the mean.

Which set of times has more variation, Scotch Pines or White Pines? →

Suppose another type of tree, Douglas fir, produced a set of times with a greater range than white pine. Does that mean this set would automatically have a greater standard deviation? Explain.

Douglas Fir Example: Answer

Even if a data set has a greater range that does not mean that it will have a greater standard deviation.
The two extremes values (the first and last numbers) may be farther away from the mean, however the remaining values might be closer or even at the mean, which would mean that the overall variation might not be greater.

See example: Douglas fir values

0.4, 10.4, 10.4, 10.4, 10.4, 10.4, 10.4, 10.4, 20.4

The range is 20.0 (20.4 - 0.4)

this is greater than for scotch and white pines

The standard deviation is 4.71

this is less than for scotch and white pines

Practice Problem

The amount of snowfall, in centimeters, on eight days were measured and recorded:

6 8 10 15 3 6 4 12

a) Find the mean snowfall.

b) Copy and complete the following table for the data:

Mean	Snowfall (Data Value)	Difference from Mean (Mean - Data value)	Squared Deviation (Difference^2)
	3		
	4		
	6		
	6		
	8		
	10		
	12		
	15		

c) Calculate the mean of the squared difference.

d) Find the square root of your answer in (c).

What is the standard deviation of the data?

Classwork/Homework

The hours of TV watched per week by 12 families in Newcastle are:

34 22 28 0 17 30 18 26 28 8 3 24

a) Find the mean of the data?

b) Copy and complete the following table for the data:

Mean	TV Watched (Data value)	Difference from Mean (Mean - Data value)	Squared deviation (Difference^2)

c) Calculate the mean of the squared difference.

d) Find the square root of your answer in (c).

What is the standard deviation of the data?